Supplementary materials

Appendices

A.1 Identifying the posterior precision matrix

To identify the precision matrix of the joint posterior distribution of $\boldsymbol{\beta}^w$ and $\boldsymbol{\theta}$ for the PCP we write:

$$\begin{aligned}
\pi(\boldsymbol{\beta}^w, \boldsymbol{\theta}|\boldsymbol{y}) &\propto \pi(\boldsymbol{Y}|\boldsymbol{\beta}^w, \boldsymbol{\theta})\pi(\boldsymbol{\beta}^w|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\
&\propto \exp\bigg(-\frac{1}{2}\Big[\big\{\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}^w - \boldsymbol{X}_1(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{X}_2\big\}^{\mathrm{T}}\boldsymbol{C}_1^{-1}\big\{\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{\beta}^w \\
&\quad -\boldsymbol{X}_1(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{X}_2\big\} + \big(\boldsymbol{\beta}^w - \boldsymbol{W}\boldsymbol{X}_2\boldsymbol{\theta}\big)^{\mathrm{T}}\boldsymbol{C}_2^{-1}\big(\boldsymbol{\beta}^w - \boldsymbol{W}\boldsymbol{X}_2\boldsymbol{\theta}\big) \\
&\quad +(\boldsymbol{\theta} - \boldsymbol{m})^{\mathrm{T}}\boldsymbol{C}_3^{-1}(\boldsymbol{\theta} - \boldsymbol{m})\Big]\bigg) \\
&= \exp\bigg(-\frac{1}{2}\Big[\ldots + \boldsymbol{\beta}^{w\mathrm{T}}\big(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1 + \boldsymbol{C}_2^{-1}\big)\boldsymbol{\beta}^w \\
&\quad +2\boldsymbol{\beta}^{w\mathrm{T}}\big\{\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{X}_2 - \boldsymbol{C}_2^{-1}\boldsymbol{W}\boldsymbol{X}_2\big\}\boldsymbol{\theta} + \boldsymbol{\theta}^{\mathrm{T}}\big\{\boldsymbol{X}_2^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{W})^{\mathrm{T}} \\
&\quad \boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{X}_2 + \boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{C}_2^{-1}\boldsymbol{W}\boldsymbol{X}_2 + \boldsymbol{C}_3^{-1}\big\}\boldsymbol{\theta} + \ldots\Big]\bigg).
\end{aligned}$$

The entries of the precision matrix can then be read off of the final expression.

A.2 Convergence rate of the PCP

Consider $\boldsymbol{Q}^{pc}_{\beta^w\theta}$ and substitute $\boldsymbol{W}$ from equation (8), then we have

$$\begin{aligned}
\boldsymbol{Q}^{pc}_{\beta^w\theta} &= \boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{X}_2 - \boldsymbol{C}_2^{-1}\boldsymbol{W}\boldsymbol{X}_2 \\
&= (\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1)\Big\{(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1 + \boldsymbol{C}_2^{-1})^{-1}\boldsymbol{C}_2^{-1}\Big\}\boldsymbol{X}_2 - \boldsymbol{C}_2^{-1}\Big\{\boldsymbol{I} - (\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1 \\
&\quad +\boldsymbol{C}_2^{-1})^{-1}\boldsymbol{C}_2^{-1}\Big\}\boldsymbol{X}_2 \\
&= \Big\{(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1)(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1 + \boldsymbol{C}_2^{-1})^{-1}\boldsymbol{C}_2^{-1} + \boldsymbol{C}_2^{-1}(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1 + \boldsymbol{C}_2^{-1})^{-1}\boldsymbol{C}_2^{-1} \\
&\quad -\boldsymbol{C}_2^{-1}\Big\}\boldsymbol{X}_2 \\
&= \Big\{(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1 + \boldsymbol{C}_2^{-1})(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1 + \boldsymbol{C}_2^{-1})^{-1}\boldsymbol{C}_2^{-1} - \boldsymbol{C}_2^{-1}\Big\}\boldsymbol{X}_2 \\
&= \Big\{\boldsymbol{C}_2^{-1} - \boldsymbol{C}_2^{-1}\Big\}\boldsymbol{X}_2 \\
&= \boldsymbol{0}.
\end{aligned}$$

Therefore by setting $\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{B}\boldsymbol{C}_2^{-1}$, $\boldsymbol{F}_{22}^{pc}$ becomes the null matrix and immediate convergence follows.

A.3 Convergence rate of the PCP

We now look at the implication of setting $\boldsymbol{W}$ according to (8) for the convergence rate of a Gibbs sampler using the PCP. For Gibbs samplers with Gaussian target distributions with known precision matrices we have analytical results for the exact convergence rate (Roberts and Sahu, 1997, Theorem 1). Convergence here is defined in terms of how rapidly the expectations of square integrable functions approach their stationary values.

Suppose that $\boldsymbol{\xi} \mid \boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We let $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ denote the posterior precision matrix. To compute the convergence rate first partition $\boldsymbol{Q}$ according to a number of blocks, denoted by l, that are used for updating $\boldsymbol{\xi}$, i.e.,

$$(\boldsymbol{Q})_{ij} = \boldsymbol{Q}_{ij}, \text{ for } i, j = 1, \ldots, l. \tag{14}$$

Let $\boldsymbol{A} = \boldsymbol{I} - \mathrm{diag}(\boldsymbol{Q}_{11}^{-1}, \ldots, \boldsymbol{Q}_{ll}^{-1})\boldsymbol{Q}$ and $\boldsymbol{F} = (\boldsymbol{I} - \boldsymbol{L}_A)^{-1}\boldsymbol{U}_A$, where $\boldsymbol{L}_A$ is the block lower triangular matrix of $\boldsymbol{A}$, and $\boldsymbol{U}_A = \boldsymbol{A} - \boldsymbol{L}_A$. Roberts and Sahu (1997) show that the Markov chain induced by the Gibbs sampler with components block updated according to matrix (14), has a Gaussian transition density with mean $E\{\boldsymbol{\xi}^{(t+1)} | \boldsymbol{\xi}^{(t)}\} = \boldsymbol{F}\boldsymbol{\xi}^{(t)} + \boldsymbol{f}$, where $\boldsymbol{f} = (\boldsymbol{I} - \boldsymbol{F})\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} - \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^{\mathrm{T}}$. Their observation leads to the following:

Theorem A.1 *(Roberts and Sahu, 1997) A Markov chain with transition density*

$$N\{\boldsymbol{F}\boldsymbol{\xi}^{(t)} + \boldsymbol{f}, \boldsymbol{\Sigma} - \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^{\mathrm{T}}\},$$

has a convergence rate equal to the maximum modulus eigenvalue of $\boldsymbol{F}$.

Corollary A.2 *If we update $\boldsymbol{\xi}$ in two blocks so that $l = 2$ then*

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{pmatrix}, \quad \boldsymbol{F} = \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12} \\ \boldsymbol{0} & \boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12} \end{pmatrix},$$

and the convergence rate is the maximum modulus eigenvalue of $\boldsymbol{F}_{22} = \boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}$.

To compute the convergence rate of the PCP we first need the posterior precision matrix of $\boldsymbol{\beta}^w$ and $\boldsymbol{\theta}$, which we can identify by writing down $\pi(\boldsymbol{\beta}^w, \boldsymbol{\theta}|\boldsymbol{y})$ explicitly (more details are provided in Appendix A.1 in Supplementary materials). The posterior precision matrix for the PCP is

$$\boldsymbol{Q}^{pc} = \begin{pmatrix} \boldsymbol{Q}^{pc}_{\beta^w} & \boldsymbol{Q}^{pc}_{\beta^w\theta} \\ \boldsymbol{Q}^{pc}_{\theta\beta^w} & \boldsymbol{Q}^{pc}_{\theta} \end{pmatrix}, \tag{15}$$

where $\boldsymbol{Q}^{pc}_{\beta^w} = \boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1 + \boldsymbol{C}_2^{-1}$, $\boldsymbol{Q}^{pc}_{\beta^w\theta} = \boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_2 - \boldsymbol{C}_2^{-1}\boldsymbol{W}\boldsymbol{X}_2$, and $\boldsymbol{Q}^{pc}_{\theta} = \boldsymbol{X}_2^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{W})^{\mathrm{T}}\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{C}_1^{-1}\boldsymbol{X}_1(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_2 + \boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{C}_2^{-1}\boldsymbol{W}\boldsymbol{X}_2 + \boldsymbol{C}_3^{-1}$. If we block update a Gibbs sampler according to the partitioning of the precision matrix (15), by Corollary A.2, we have that the convergence rate of the PCP is the maximum modulus eigenvalue of the matrix $\boldsymbol{F}^{pc}_{22} = (\boldsymbol{Q}^{pc}_{\theta})^{-1}\boldsymbol{Q}^{pc}_{\theta\beta^w}(\boldsymbol{Q}^{pc}_{\beta^w})^{-1}\boldsymbol{Q}^{pc}_{\beta^w\theta}$. By construction we have a 2×2 block diagonal posterior covariance matrix for $\boldsymbol{\beta}^w$ and $\boldsymbol{\theta}$. Therefore the precision matrix is also block diagonal and $\boldsymbol{F}^{pc}_{22}$ is null and immediate convergence is achieved.

A.4 Convergence rate of a three component Gibbs sampler

It can be shown that a Gibbs sampler with Gaussian target distribution with precision matrix given by $\boldsymbol{Q}$ having elements $(\boldsymbol{Q})_{ij} = \boldsymbol{Q}_{ij}$ for $i, j = 1, 2, 3$ has a convergence rate which is equal to the maximum modulus eigenvalue of

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12} & -\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{13} \\ \boldsymbol{0} & \boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12} & \boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{13} - \boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{23} \\ \boldsymbol{0} & \boldsymbol{F}_{32} & \boldsymbol{F}_{33} \end{pmatrix},$$

where

$$\begin{aligned} \boldsymbol{F}_{32} &= (\boldsymbol{Q}_{33}^{-1}\boldsymbol{Q}_{31} - \boldsymbol{Q}_{33}^{-1}\boldsymbol{Q}_{32}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21})\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}, \\ \boldsymbol{F}_{33} &= (\boldsymbol{Q}_{33}^{-1}\boldsymbol{Q}_{31} - \boldsymbol{Q}_{33}^{-1}\boldsymbol{Q}_{32}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21})\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{13} + \boldsymbol{Q}_{33}^{-1}\boldsymbol{Q}_{32}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{23}. \end{aligned}$$

A.5 Proof of stationarity of the PCP

To demonstrate that stationarity is preserved we let $p = 1$ in model (4). The transition kernel of the Markov chain is:

$$\begin{aligned} P\{\boldsymbol{\xi}^{(t+1)}|\boldsymbol{\xi}^{(t)}) &= \pi\{\boldsymbol{\beta}^{w(t+1)}|\theta_0^{(t)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\}\pi\{\theta_0^{(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\} \\ &\quad \pi\{\sigma_0^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\} \\ &\quad \pi\{\sigma_\epsilon^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\}. \end{aligned}$$

We have dropped the $\boldsymbol{W}$'s to save space, conditioning the variance parameters on their current values where necessary. It follows that

$$\begin{aligned} &\int P\{\boldsymbol{\xi}^{(t+1)}|\boldsymbol{\xi}^{(t)}\}\pi(\boldsymbol{\xi}^{(t)}|\boldsymbol{y}\}d\boldsymbol{\xi}^{(t)} \\ &= \int \pi\{\sigma_\epsilon^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\} \\ &\quad \pi\{\sigma_0^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\}\pi\{\theta_0^{(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\}d\boldsymbol{\xi}^{(t)} \end{aligned}$$

$$\begin{aligned} &= \int \pi\{\sigma_\epsilon^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\}\pi\{\sigma_0^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\} \\ &\quad \underbrace{\pi\{\theta_0^{(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\}\left[\int \pi\{\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}|\boldsymbol{y}\}d\theta_0^{(t)}\right]}_{=\pi\{\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}|\boldsymbol{y}\}} d\sigma_0^{2(t)} d\sigma_\epsilon^{2(t)} \\ &= \int \pi\{\sigma_\epsilon^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\} \\ &\quad \underbrace{\left[\int \pi\{\sigma_0^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\}\pi\{\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}|\boldsymbol{y}\}d\sigma_0^{2(t)}\right]}_{=\pi\{\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t)}|\boldsymbol{y}\}} d\sigma_\epsilon^{2(t)} \\ &= \int \pi\{\sigma_\epsilon^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\}\pi\{\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t)}|\boldsymbol{y}\}d\sigma_\epsilon^{2(t)} \\ &= \pi\{\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t+1)}|\boldsymbol{y}\} \\ &= \pi\{\boldsymbol{\xi}^{(t+1)}|\boldsymbol{y}\}, \end{aligned}$$

and hence stationarity is preserved. The above argument can easily be extended for $p > 1$ or to include other correlation parameters if they are being modelled.

If we update $\boldsymbol{W}$ and the end of each complete pass of the sampler then the stationarity condition (10) does not hold. For instance, consider σ_ϵ^2, which is conditioned on σ_0^2 through $\boldsymbol{W}$. If $\boldsymbol{W}$ is not recalculated using $\sigma_0^{2(t+1)}$ then $\sigma_\epsilon^{2(t+1)}$ is conditioned and $\sigma_0^{2(t)}$, and consequently

$$\begin{aligned}&\int \pi\{\sigma_\epsilon^{2(t+1)}|\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\}\pi\{\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t)}|\boldsymbol{y}\}d\sigma_\epsilon^{2(t)} \\ &\neq \pi\{\boldsymbol{\beta}_0^{w(t+1)}, \theta_0^{(t+1)}, \sigma_0^{2(t+1)}, \sigma_\epsilon^{2(t+1)}|\boldsymbol{y}\},\end{aligned}$$

but equality is required to complete step (16) in the string of equalities proving stationarity.

A.6 Joint posterior and full conditional distributions

We begin here by writing down the joint posterior distribution of the parameters in model (4). We let $\boldsymbol{\xi} = (\boldsymbol{\beta}^{w\mathrm{T}}, \boldsymbol{\theta}^{\mathrm{T}}, \boldsymbol{\sigma^2}^{\mathrm{T}}, \sigma_\epsilon^2)^{\mathrm{T}}$ be the vector containing all np partially centred random effects, p global effects, p random effect variances, the data variance and p decay parameters for the correlation functions. The joint posterior for $\boldsymbol{\xi}$ is

$$\begin{aligned}\pi(\boldsymbol{\xi}|\boldsymbol{y}) &\propto \pi(\boldsymbol{Y}|\boldsymbol{\beta}^w, \boldsymbol{\theta}, \sigma_\epsilon^2)\pi(\boldsymbol{\beta}^w|\boldsymbol{\theta}, \boldsymbol{\sigma^2})\pi(\boldsymbol{\theta}|\boldsymbol{\sigma^2})\pi(\boldsymbol{\sigma^2})\pi(\sigma_\epsilon^2) \\ &\propto \prod_{k=0}^{p-1}\left(\sigma_k^2\right)^{-(n/2+1/2+a_k+1)}|\boldsymbol{R}_k|^{-1/2}\left(\sigma_\epsilon^2\right)^{-(n/2+a_\epsilon+1)} \\ &\quad \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\Big[\boldsymbol{Y} - \boldsymbol{X}_1\{\boldsymbol{\beta}^w + (\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_2\boldsymbol{\theta}\}\Big]^{\mathrm{T}}\Big[\boldsymbol{Y} - \boldsymbol{X}_1\{\boldsymbol{\beta}^w + \right.\right. \\ &\left.\left. (\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_2\boldsymbol{\theta}\}\Big] + 2b_\epsilon\right)\right\}\exp\left\{-\frac{1}{2}\Big(\boldsymbol{\beta}^w - \boldsymbol{W}\boldsymbol{X}_2\boldsymbol{\theta}\Big)^{\mathrm{T}}\boldsymbol{C}_2^{-1}\Big(\boldsymbol{\beta}^w - \boldsymbol{W}\boldsymbol{X}_2\boldsymbol{\theta}\Big)\right\} \\ &\quad \exp\left[-\frac{1}{2}\sum_{k=0}^{p-1}\frac{1}{\sigma_k^2}\left\{\frac{(\theta_k - m_k)^2}{v_k} + 2b_k\right\}\right],\end{aligned}$$

where a description of the prior distributions $\pi(\boldsymbol{\sigma^2})$ and $\pi(\sigma_\epsilon^2)$ can be found in Section 2.1.

It is argued in Section A.3 that we must jointly update the $\boldsymbol{\beta}^w$'s and jointly update $\boldsymbol{\theta}$ and this is reflected in the conditional distributions given below.

- The full conditional distribution of $\boldsymbol{\beta}^w$ is $\boldsymbol{\beta}^w|\boldsymbol{\theta}, \boldsymbol{\sigma^2}, \sigma_\epsilon^2, \boldsymbol{y} \sim N\left(\boldsymbol{m}_\beta^*, \boldsymbol{C}_2^*\right)$, where $\boldsymbol{C}_2^* = \left(\sigma_\epsilon^{-2}\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1 + \boldsymbol{C}_2^{-1}\right)^{-1}$ and $\boldsymbol{m}_\beta^* = \boldsymbol{C}_2^*\left[\sigma_\epsilon^{-2}\{\boldsymbol{y} - \boldsymbol{X}_1(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_2\boldsymbol{\theta}\} + \boldsymbol{C}_2^{-1}\boldsymbol{W}\boldsymbol{X}_2\boldsymbol{\theta}\right]$.

- The full conditional distribution of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}|\boldsymbol{\beta}^w, \boldsymbol{\sigma^2}, \sigma_\epsilon^2, \boldsymbol{y} \sim N\left(\boldsymbol{m}_\theta^*, \boldsymbol{C}_3^*\right)$, where

$$\boldsymbol{C}_3^* = \left[\sigma_\epsilon^{-2}\{\boldsymbol{X}_1(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_2\}^{\mathrm{T}}\{\boldsymbol{X}_1(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_2\} + (\boldsymbol{W}\boldsymbol{X}_2)^{\mathrm{T}}\boldsymbol{C}_2^{-1}\boldsymbol{W}\boldsymbol{X}_2 + \boldsymbol{C}_3^{-1}\right]^{-1},$$
$$\boldsymbol{m}_\theta^* = \boldsymbol{C}_3^*\left[\sigma_\epsilon^{-2}\{\boldsymbol{X}_1(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_2\}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{X}_1\boldsymbol{\beta}^w) + (\boldsymbol{W}\boldsymbol{X}_2)^{\mathrm{T}}\boldsymbol{C}_2^{-1}\boldsymbol{\beta}^w + \boldsymbol{C}_3^{-1}\boldsymbol{m}\right].$$

- The full conditional distribution of σ_k^2 is $\sigma_k^2|\boldsymbol{\beta}^w, \boldsymbol{\theta}, \boldsymbol{\sigma^2}_{-k}, \sigma_\epsilon^2, \boldsymbol{y}$

$$IG\left[\frac{n+1}{2} + a_k,\ b_k + \frac{1}{2}\left\{\left(\boldsymbol{\beta}_k^w - \boldsymbol{\eta}_k\right)^{\mathrm{T}} \boldsymbol{R}_k^{-1}\left(\boldsymbol{\beta}_k^w - \boldsymbol{\eta}_k\right) + \frac{(\theta_k - m_k)^2}{v_k}\right\}\right],$$

  for $k = 0, \ldots, p-1$, where $\boldsymbol{W}_{km}$ denotes the kmth, $n \times n$ block of $\boldsymbol{W}$ and $\boldsymbol{\eta}_k = \theta_k \sum_{m=0}^{p-1} \boldsymbol{W}_{km}\boldsymbol{1}$.

- The full conditional distribution of σ_ϵ^2 given $\boldsymbol{\beta}^w, \boldsymbol{\theta}, \boldsymbol{\sigma^2}, \boldsymbol{y}$ is

$$IG\left[\frac{n}{2} + a_\epsilon,\ \ b_\epsilon + \frac{1}{2}\left\{\left(\boldsymbol{Y}-\boldsymbol{Z}\right)^{\mathrm{T}}\left(\boldsymbol{Y}-\boldsymbol{Z}\right)\right\}\right],$$

  where $\boldsymbol{Z} = \boldsymbol{X}_1\left(\boldsymbol{\beta}^w + (\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_2\boldsymbol{\theta}\right)$.